This is a draft of a chapter. The chapter has not yet been through the editorial process and is not yet published. The chapter of record will be available in:

Neal, T.M.S., Hight, M., Howatt, B., & Hamza, C. (in press). The cognitive and social psychological bases of bias in forensic mental health judgments. In M.K. Miller & B.H. Bornstein (Eds). Advances in Psychology and Law: Volume 3. New York: Springer.

The Cognitive and Social Psychological Bases of Bias in Forensic Mental Health Judgments

Tess M.S. Neal, Morgan Hight, Brian Howatt, & Cassandra Hamza

Arizona State University

Author Note

Tess M.S. Neal, Morgan Hight, Brian Howatt, and Cassandra Hamza, School of Social and Behavioral Sciences, New College of Interdisciplinary Arts & Sciences, Arizona State University.

Correspondence concerning this chapter should be addressed to Tess M.S. Neal, New College of Interdisciplinary Arts & Sciences - SBS, Arizona State University, 4701 West Thunderbird Rd, Mail Code 3051, Glendale AZ 85306. E-mail: Tess.Neal@asu.edu

Abstract

This chapter integrates from cognitive neuroscience, cognitive psychology, and social psychology the basic science of bias in human judgment as relevant to judgments and decisions by forensic mental health professionals. Forensic mental health professionals help courts make decisions in cases when some question of psychology is relevant to the legal issue, such as in insanity cases, child custody hearings, and psychological injuries in civil suits. The legal system itself and many people involved, such as jurors, assume mental health experts are "objective" and untainted by bias. However, basic psychological science from several branches of the discipline suggest the law's assumption about experts' protection from bias is wrong. Indeed, several empirical studies now show clear evidence of (unintentional) bias in forensic mental experts' judgments and decisions. In this chapter, we explain the science of how and why human judgments are susceptible to various kinds of bias. We describe dual-process theories from cognitive neuroscience, cognitive psychology, and social psychology that can help explain biased judgments. We review the empirical evidence to date specifically about cognitive and social psychological biases in forensic mental health judgments, weaving in related literature about biases in other types of expert judgment with hypotheses about how forensic experts are likely affected by these biases. We close with a discussion of directions for future research and practice.

Keywords: judgment; decision; bias; forensic; dual-process; cognitive; social; heuristic; implicit

The Cognitive and Social Psychological Bases of Bias in Forensic Mental Health Judgments

This chapter reviews the basic psychological science of bias in human judgment as relevant to judgments and decisions by forensic mental health professionals. The legal system itself and many people involved, such as jurors, assume mental health experts can be and typically are "objective" and protected by bias. Many experts themselves believe they can control their various biases in order to practice objectively. Indeed, psychology ethics codes and guidelines require that practicing psychologists be objective. However, basic psychological science from several branches of the discipline suggest these assumptions about experts' protection from bias is wrong. Empirical studies now show clear evidence of (unintentional) bias in forensic mental experts' judgments and decisions. This chapter explains how and why human judgments are susceptible to various kinds of bias, with specific emphasis on expert judgments, particularly but not exclusively in the domain of forensic psychology. The implications across these findings for bias mitigation are discussed, as are promising directions for bias mitigation. We close with a discussion of directions for future research and practice.

The Psychology of Bias in Cognitive and Social Judgments

Much of this chapter will focus on cognitive psychological issues of mental functioning, such as perception, reasoning, attention, memory, and decision making with coverage of core cognitive psychological literature. However, this chapter will also focus on social psychological issues (particularly, social cognition) involving how other people affect our perceptions, reasoning, attention, memory, and decision making. Some areas in which social cognition can inform the topics discussed in this chapter include attitudes, stereotypes, impression formation, the self, person perception, attribution, persuasion, and conformity (Chaiken & Trope, 1999;

Devine, Hamilton, & Ostrom, 1994; Greenwald & Banaji, 1995). The dual-process theoretical foundation of this chapter is inherent in both the cognitive and social-cognitive literatures.

Human cognitive processing abilities are in many ways extraordinary. Our cognitive processing machinery has evolved into an "adaptive toolbox" to help us efficiently process the vast amounts of information we're faced with each day (e.g., Hoffrage & Gigerenzer, 2004). We literally wouldn't be able to function if we didn't have "shortcuts" that work well most of the time to help us reduce the complexities of our environment, cope with information overload, and make reasonable judgments and decisions. Haselton, Bryant, Wilke, Frederick, & Galperin (2009) and Gigerenzer (2008) describe how our adaptively rational mind developed based on our ancestors' need for survival, and how the influence of mechanisms from the past influence our actions today.

But this very design also makes us susceptible to making predictable and systematic errors (see e.g., Kahneman, 2011; Kahneman & Klein, 2009). For example, our prior expectations distort our subsequent processing of new information. Our personal self-interest shapes our interpretation of "objective" data and facts. Other people affect our judgments in ways we are unaware. Even random numbers in our environment can act as "anchors" that pull our numeric estimates away from the truth and toward something irrelevant. These biases, one might hope and expect, should be less likely to impact experts—particularly experts making judgments and decisions in their domain of expertise – like forensic psychologists in forensic psychological evaluations. Yet, the evidence suggests that experts are indeed human – and that their brains work like the rest of ours, for better and for worse. The types of biases we focus on in this chapter are unintentional, "System 1" biases that are below the level of conscious

awareness, as we explain next. Although experts can be biased by other types of more intentional and explicit biases, this chapter focuses on unintentional biases.

Dual-Process Theories

Evidence from cognitive neuroscience, cognitive psychology, and social-cognition is converging on the conclusion that human brain functioning can be characterized by two types of cognition, each with different functions and strengths and weaknesses (Chaiken & Trope, 1999; Greenwald & Banaji, 1995; Sloman, 1996; Stanovich & West, 2000; Kahneman, 2003). System 1 (sometimes referred to as Type 1) processing occurs automatically, quickly, with little or no effort, and implicitly – that is, below the level of conscious awareness. In contrast, System 2 (Type 2) processing is slow, deliberate, effortful, and explicitly conscious. These distinctions are important, because in this chapter we focus on System 1 biases that occur quickly, automatically, and outside of the awareness of experts. Experts are human, and thus it is reasonable and even expected that they should experience the kinds of cognitive strengths and limitations that other humans face. Understanding these factors can inform how to better structure the decision tasks experts face in order to minimize the chances that System 1 errors and biases can affect experts' judgments.

While the preceding research has revealed the qualities of dual-*processes* of cognition within a dual-*system* framework, some researchers have argued against this notion. Evans' (2008) review of dual-system models led him to suggest that dual-systems do not necessarily follow from dual-processes, calling this conclusion "oversimplified and misleading" (p. 270). He argues the data show the processes are not mutually exclusive, either structurally or functionally, and that they each share conscious and preconscious operations. Instead, he proposes that

reflective cognition requires access to a central working-memory repository, in which overload or interference can then disrupt its processes – an interactive approach to cognition.

Cognitive Psychology: The Heuristics and Biases Literature

Some researchers have argued that System 1-induced cognitive "biases" are not really biases (Gigerenzer, 1991), but rather illustrate an adaptive rationality (Gigerenzer, 2008; Haselton et al., 2009) that aids an organism's survival by rapidly evaluating environmental data and guiding appropriate behavior. However, these swift processes are nonetheless prone to error in systematic ways. Thus, "bias," as used in the heuristics and biases tradition of cognitive psychology, is a by-product of System 1 mental shortcuts. The framing of System 1 processes as problematic "biases" versus "adaptive strengths" is a longstanding point of contention in the literature (see e.g., Kahneman & Klein, 2009), which is relevant for thinking about experts' decisions processes. Although there is disagreement about the types of heuristics and how they are observed, there seems to be a general consensus regarding heuristics as highly serviceable aspects of cognition, though subject to error.

System 1 "heuristics," or mental shortcuts that expedite the time and effort required to process and interpret information, involve a speed/accuracy tradeoff allows an individual to make judgments and decisions that are "good enough" even though a more maximal outcome may exist. Notable heuristics initially developed by decision researchers include *representativeness* (Kahneman & Tversky, 1972), *availability* (Tversky & Kahneman, 1973), and *anchoring* (Tversky & Kahneman, 1974), with the later addition of *affect* (Slovic, Finucane, Peters, & MacGregor, 2007). Representativeness is a way of organizing information based upon its similarity to other information, whereas availability affects judgments based on how easy or it is to recall other examples of an event in question. Anchoring affects judgments in that initial

information encountered is more heavily weighted than subsequent information. Affect is an emotion-oriented heuristic, in which people pursue pleasure and avoid pain. While these methods are generally sufficient in cuing appropriate behavior, they can distort the likelihood or magnitude of an event's occurrence, leading to systematic biases in judgments and choices.

Gigerenzer (1991, 1996) has argued against these specific heuristics, calling them merely labels of behavior, and devoid of ecological applicability. Alternatively, he has put forth common misconceptions of heuristics and proposed an "adaptive toolbox" (Gigerenzer, 2008) of behavioral processes, including satisficing, recognition, and default. Satisficing is choosing the first option that meets an acceptable level of criteria; the recognition heuristic places a greater value on options that are recognizable; and the default heuristic strives to maintain the status quo, unless another option is far superior. Moreover, he claims these adaptive heuristics can outperform objective, mathematical analyses of decision scenarios (e.g., System 2 processes) and calls for research to focus more heavily on the interaction between these strategies and the environments in which they are made. Haselton and colleagues' (2009) research seems to substantiate Gigerenzer's proposals, suggesting heuristics are not as flawed as previous research posits and that limited environmental information may lead to erroneous conclusions.

It should be noted that Kahneman and Tversky acknowledged the efficacy of heuristics in daily judgments and decisions, but were more interested in examining how resulting behavior deviated from theories postulating the optimality of statistical intuition (which Gigerenzer also acknowledges can readily happen). And if such aberrations were not merely trace artifacts, but systematic, reproducible, and predictable violations, then it should be possible to implement corrective measures.

In many ways, experts develop ways of recognizing and diagnosing situations that non-experts cannot – by virtue of both System 1 and System 2 processes. But System 1 can sometimes lead experts astray, even in their own domains of expertise. The reason it's important to recognize how and when this can happen – to identify some common problems – is that we can then develop methodologies to ameliorate their negative influences.

Cognitive Psychological Biases and Forensic Mental Health Judgments

This section considers various cognitive biases that may affect forensic mental health judgments. We also discuss the role of bias in judgments and decision making in similar expert judgment situation outside of the forensic mental health field in order to expand this discussion. Should the reader be interested in some thought experiments and examples of how the heuristics mentioned above might affect forensic mental health judgments, see Neal and Grisso (2014).

For example, Neal and Grisso (2014) provide a vignette about "John P." and his potential mental illness at the time of an alleged crime to illustrate the representativeness heuristic as relevant to forensic mental health, as well as a related real-world example of an internationally recognized forensic psychiatrist neglecting relevant base-rates in the John Hinckley trial (1982 trial of the attempted assassination of President Ronald Reagan). To illustrate the availability heuristic, Neal and Grisso discuss the problem of false negatives in sex offender risk assessments, invite readers to attempt an adapted version of the Wason (1968) card task, and describe the relevance of Kahneman's (2011) WYSIATI (What You See Is All There Is) concept for forensic mental health assessments. WYSIATI is closely related to the availability heuristic: only ideas that are activated in a person's mind are processed within a given decision task. Finally, to illustrate anchoring, Neal and Grisso describe a case in which one might encounter one likeable parent versus the other in a child custody evaluation and they further describe

framing and context effects as relevant to forensic decision tasks, to be discussed in more depth in the following section.

Evidence to date specifically relevant to forensic mental health judgments. Even for experts who are motivated to be unbiased, there is mounting evidence that forensic mental health experts are susceptible to System 1-induced biases by virtue of being human. For example, clear evidence of the "self-serving bias," which has been labeled "adversarial allegiance" in the forensic mental health field has been documented. Adversarial allegiance is an unintentional tendency for experts to find evidence in support of their retaining party's position – an anchoring-like effect – and has been uncovered in forensic mental health judgments (Murrie, Boccaccini, Guarnera, & Rufino, 2013). Confirmation bias, or a tendency for experts to seek evidence in support of an initial hypothesis without seeking disconfirming evidence – another type of System 1 bias – has also been documented in forensic mental health experts' judgments (Neal, MacLean, Morgan, & Murrie, 2017) And hindsight bias, another System 1 bias, has been found in forensic psychiatrists' judgments (LeBourgeois, Pinals, Williams, & Appelbaum, 2007).

Self-Serving Bias. Affiliations with other people affect people's processing of information. One of the first studies of the power of affiliation documenting the self-serving bias was actually with regard to sports teams. In a classic study, Hastdorf and Cantril (1954) showed that football fans from each team blamed the other team for behaving badly while discounting their own team's behaviors. In a more legally-relevant study, Babcock and Loewenstein (1997) showed that even when provided with the same case information, people on opposing sides of a case reach different conclusions based on that evidence – in favor of their own interests. In this study, Babcock and Loewenstein provided pairs of participants with police and medical reports, depositions, and other materials from a lawsuit resulting from a collision between a motorcycle

and a car. Participants were randomly assigned to the role of either the motorcyclist plaintiff or car-driving defendant. Participants in the motorcyclist plaintiff role found evidence to support their positon and predicted dramatically larger damage awards than the defendants.

In the American adversarial legal system, forensic mental health experts are often retained by one side or another in a case. Involvement and affiliation with attorneys has been shown to influence experts' examination and evaluation of case materials in that experts end up emphasizing findings and patterns that support "their side." Murrie, Boccaccini and colleagues documented this phenomenon in observational studies where they found clear patterns of experts scoring standardized psychological tools in favor of their retaining party in sexual offender civil commitment cases (Murrie, Boccaccini, Johnson, & Janke, 2008; Murrie, Boccaccini, Turner, Meeks, Woods, & Tussey, 2009). For example, Murrie et al. (2009) looked at real sexual offender civil commitment cases and measured the extent to which experts on each "side" of the case scored three tools that are often used in sex offender risk assessments: the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R; Epperson et al., 1998), Psychopathy Checklist-Revised (PCL-R; Hare, 2003), and the STATIC-99 (Hanson & Thornton, 1999).

Consistent with the self-serving bias / adversarial allegiance hypothesis, Murrie et al. (2009) showed that plaintiff-retained experts scored the offenders on average 3.53 points higher than respondent-retained experts on the MnSOST-R (Cohen's d = 0.85, a large effect size), 5.79 points higher on the PCL-R (Cohen's d = 0.78, a large effect size), and 0.52 points higher on the STATIC-99 (Cohen's d = 0.34, a moderately small effect size on a measure with less subjective items). Furthermore, they calculated the proportion of variance in the scores on these tools that were attributable to the offender (which ideally would be 100%), the side of the case (ideally 0%), and other error (ideally 0%). They found that 44% of the variance on the MnSOST-R was

attributable to the offender, 26% was attributable to the side of the case (this is the adversarial allegiance effect), and 30% was other error. On the PCL-R, 42% of the variance was attributable to the offender, 23% to the side of the case (adversarial allegiance), and 35% to other error. And for the most structured and least subjective measure, the STATIC-99, 62% of the variance was attributable to the offender, only 4% to the side of the case (adversarial allegiance), and 34% to other error.

The systematic evidence of adversarial allegiance in forensic mental health evaluations led researchers to probe more deeply into potential explanations for the findings. Were forensic mental health evaluators intentionally biased "hired guns?" Or were they unintentionally biased and unaware of it – biased by unconscious System 1 processes by virtue of being human? Was it the mere fact of the adversarial hire itself that caused the adversarial allegiance bias, or was it something else (e.g., self-selection factors, like experts choosing which side to work for based on pre-existing biases)? Murrie et al. (2013) designed an elegant experiment to explore some of these questions.

In an experimental study of civil commitment proceedings for sex offenders, Murrie et al., (2013) had real forensic mental health experts hired by a referring attorney via a script (randomly assigned to either plaintiff or respondent). That is, the manipulated independent variable was the adversarial "side" for which experts thought they were working (the case and offender materials were held constant). The participant-experts were asked to score four offenders on two commonly-used and well-researched risk instruments, the PCL-R and the STATIC-99. Importantly, the forensic evaluators were paid $400 for their consultation and were deceived to believe these referrals were real, as opposed to being part of a research study: they did not know they were being studied.

Results revealed a significant bias as a function of the side by which the expert was retained: more evidence of adversarial allegiance. Forensic mental health experts who believed they were working for the plaintiff assigned higher scores on the risk instruments than experts who believed they were working for the respondent. The effect sizes were up to $d = 0.85$ (large effects) for the PCL-R and up to $d = 0.42$ on the STATIC-99 (moderately small effect on this measure with less subjective items). Of course there is considerable variability in scoring assessment instruments: not every evaluator will produce consistent scores. But here, a large portion of the systematic score differences amongst opposing experts was explained by adversarial allegiance and not by chance or random error.

The study demonstrated that experts scoring ostensibly objective assessment instruments assigned scores that were systematically biased towards the side that retained them. Given the experimental design of the study, cause can be inferred: the only variable that differed between the two conditions was the hiring party. Participants were unaware that the hiring party influenced their scores, yet there was clear evidence that adversarial allegiance absolutely influenced experts' scores. Thus, Murrie and colleagues (2013) attributed the adversarial allegiance effect directly to experts' beliefs about for whom they were working, because they controlled for other possible explanations.

The substantive information provided about the offender was constant, so differences in the way the examinee presented could not have explained the findings. Furthermore, they eliminated the overt verbal influence often provided by the referral party in routine forensic practice that contributes to confirmation bias by using a script. This design element is important: their findings show that even when there is no overt framing by a referral party, there is still an

insidious, unconscious, and potent form of anchoring due to adversarial allegiance affecting forensic mental health professionals' judgments and decisions.

McAuliff & Arter (2016) studied the potential influence of adversarial allegiance on different aspects of expert testimony in simulated child sexual abuse case. Participant experts were asked by either the prosecution or the defense to read a description of a police officer's low or high suggestive interview with a 5-year-old girl. Experts were more willing to testify if asked by the prosecution when the suggestibility in the police officer interview was low, meaning when the interview was done soundly by the officer and did not raise concerns about the child's accuracy. Whereas the experts that were asked by the defense were more willing to testify when the suggestibility of the interview was high, meaning the interview between the police officer was unsound and concerns about the child's accuracy were raised. Thus, experts may have perceived their testimony as being more relevant and more helpful to jurors when the evidence favored the party soliciting their testimony – more evidence suggestive of adversarial allegiance and how the adversarial system influences forensic mental health professionals' judgments and decisions.

Confirmation bias. Neal et al. (2017) conducted a recent study of confirmation bias in forensic psychologists' diagnostic reasoning. Confirmation bias is a System-1 type of process in which people tend to seek and rely on information that confirms a "hunch" rather than seeking disconfirmatory information (see Nickerson, 1998). The modern scientific method evolved in part to combat the powerful "confirmatory" bias in hypothesis-testing (Popper, 1959; Neal & Saks, in preparation), yet evidence of confirmation bias persists in many contexts. Some examples include intelligence analysis (Cook & Smallman, 2008), criminal investigations (Ask

& Granhag, 2005), radiology diagnostic tasks (Drew, Vo, & Wolfe, 2013), and even in science itself (MacCoun, 1998).

A national sample of 118 randomly-selected experienced forensic psychologists were invited via an email invitation to participate in the study (17% response rate). Participants were asked to provide diagnostic hypotheses for and answer questions about one of four randomly-assigned vignettes of people presenting with different sets of symptoms and from different referral contexts. The initial diagnostic question asked participants to rank-order a list of four possible initial diagnostic hypotheses "in order of likelihood that this person may meet DSM-5 diagnostic criteria for each" (the options were the same across vignettes). From there, they received a piped follow-up question linked to the diagnostic hypothesis they rank-ordered first.

The follow-up question asked "Now, based on your primary diagnostic hypothesis that Mr. G meets criteria for x, what piece of information would you want first in order to effectively test your primary diagnostic hypothesis?" They were provided with a choice between two types of information: one that might *confirm* their initial hypothesis, and one that might *disconfirm* their initial hypothesis. For example, for participants who rank-ordered Intellectual Disability as their first diagnostic hypothesis, they had the choice between "Did Mr. G show deficits in intellectual functioning on the standardized intelligence tests he took at ages 10 and 14?" (*confirmatory*) and "Does Mr. G have a personality disorder that could explain his symptoms?" (*disconfirmatory*). We hypothesized clinicians would be more likely to choose the confirmatory than disconfirmatory information (i.e., engage in confirmation bias).

The survey also included the three-item (M=1.41, SD=1.17) Cognitive Reflection Task (Frederick 2005), which had good reliability in this sample, alpha=0.72. Cognitive reflection is the ability to reflect on a question and resist the first "heuristic" response that comes to mind,

instead engaging in deliberative thought to reach an answer. We predicted clinicians with higher cognitive reflection tendencies would be less likely to engage in confirmatory bias.

Results indicated that forensic clinicians overwhelmingly engaged in confirmation bias: 103 of the 111 people who responded to this question chose the confirmatory information, χ^2 (1) = 81.31, $p < 0.001$. Cognitive reflection had a statistically and theoretically significant association with confirmation bias in the predicted direction. Each unit higher on the three-item cognitive reflection task (representing higher cognitive reflection tendencies) halved the odds of confirmatory bias, $B = -0.75$, $Wald(1) = 3.85$, $p = 0.050$, $Exp(B) = 0.473$ (logistic regression model χ^2 [1] = 4.67, $p = 0.031$).

These findings demonstrate robust confirmation bias in diagnostic reasoning in a representative sample of licensed psychologists in forensic practice in the US. Susceptibility to this bias was related to lower cognitive reflection tendencies (i.e., tendency to rely more on System 1 than System 2 thinking). What this study does not tell us is how far confirmation bias persists (how far beyond the "first piece of information"). Neal and colleagues are currently collecting new data to explore this question.

Hindsight bias. Hindsight bias is a System 1 type of bias in which people who know the outcome of an event or situation overestimate what they could have known in foresight (Fischhoff, 1975). People who know an outcome can't unknow it – it's like unhearing a bell that was rung – and such knowledge influences people's beliefs about how predictable or foreseeable the outcome actually was. This bias is highly relevant for legal cases, because legal decision makers must reason *ex post facto* and they know the outcome of the situation (i.e., the crime in criminal cases or the tort in civil cases). This outcome knowledge has been shown to bias

decision makers, including jurors (e.g., Labine & Labine, 1996) and judges (e.g., Guthrie, Rachlinksi, & Wistrich, 2001).

In clinical contexts, previous research has shown that physicians are susceptible to hindsight bias as well (Arkes, Wortmann, Saville, & Harkness, 1981; Caplan, Posner, & Cheney, 1991; Sacchi & Cherubini, 2004). For example, in a study by Caplan et al. (1991), experienced anesthesiologists were provided with a set of clinical case scenarios with set facts and were asked to review a previous physician's decisions in the case and rate the standard of care. Importantly, they were randomly assigned to know different outcomes of the cases (i.e., temporary vs. permanent injury). Anesthesiologists who knew the outcome was a permanent injury were more likely to rate the previous physician's care as substandard as compared to knowledge that the injury was just temporary. This finding is significant because the cases themselves were identical, and the outcome of the situation could not have been known at the time of the event itself. Thus, the retrospective judgments of the appropriateness of care delivered by others physicians were but should not have been influenced by the outcome of the situation.

Extending this work into the domain of forensic mental health judgments, LeBourgeois and colleagues (2007) exposed psychiatrists to hypothetical cases in which patients with suicidal or homicidal thoughts presented for care. Psychiatrist participants were randomly assigned to different outcome knowledge conditions: either that a suicide/homicide occurred shortly after the patients were released from care (hindsight group) or no information about outcome (control group). Psychiatrist participants were asked to estimate the likelihood that suicide or homicide would occur upon the patient's release. Results revealed that forensic psychiatrists were indeed affected by hindsight bias: psychiatrists who knew the outcome of the case rated the patient as at

a significantly higher risk of suicide/homicide than those in the control condition who did not know the outcome of the case. LeBourgeois et al. did not ask participants to what degree knowledge of the outcome influenced their judgments. However, given the nature of the hindsight bias and findings from other hindsight bias studies, it is likely the psychiatrists would not have been aware of how much knowledge of the outcome actually affected their judgments, would have estimated what they would have known without the outcome knowledge, and would have overestimated what others actually did know without the outcome knowledge, consistent with a System 1-induced bias (Fischhoff, 1975).

In a recent study of hindsight bias with forensic psychologists as participants, Beltrani, Reed, Zapf, Dror, and Otto (2017) used a method similar to LeBourgeois et al. (2007) and found evidence of hindsight bias in forensic psychologists' decision processes as well. Participants provided with outcome information regarding risk assessment evaluations were more likely to indicate that they would have predicted the outcome than evaluators who were not provided with outcome information, χ^2 (1) = 4.215, p = 0.04, φ = 0.235. Furthermore, when asked to provide reasons for their decisions regarding risk, participants in the known-outcome condition provided more risk factors from the initial case information to support their decisions compared to participants in the control condition, who selected more protective factors to support their decisions. These results are consistent with motivated reasoning, a social-cognitive theory proposing that motivation can affect reasoning through biased cognitive processes regarding how information is accessed, constructed, and evaluated (Kunda, 1990). This theory holds that people use the tools of cognition to arrive at desired conclusions, constrained only by one's ability to construct reasonable justifications for that conclusion (Kunda).

Related literature on cognitive biases in similar types of judgment tasks. The existence of cognitive biases in similar types of judgment tasks but outside of forensic mental health are worth covering briefly. This is because these biases are likely relevant in forensic mental health too, and thus we might generate hypotheses from these studies about how these biases might affect the work of forensic mental health professionals. For example, the judgments and decisions of forensic scientists have been shown to be subject to the effects of various cognitive biases (e.g., Dror, Charlton, & Péron, 2006; Nakhaeizadeh, Dror, & Morgan, 2014), as have judges (Guthrie et al, 2001; Wistrich, Guthrie, & Rachlinski, 2005) elite arbiters (Helm, Wistrich, & Rachlinski, 2016), and professional accountants (e.g., Moore, Loewenstein, Tanlu, & Bazerman, 2003).

Several studies in the forensic science context have revealed the power of "context effects" or extraneous information to a case that biases expert's judgments. For instance, Dror et al. (2006) studied whether latent print identification experts were vulnerable to context effects. In an inventive method, they asked experts to make a fingerprint match determination on a set of fingerprints they had previous examined and made a positive-match decision. The experts were not aware that they were looking at prints they had previously positively identified as a match. Crucially, the experimenters provided contextual information suggesting the prints were a no-match. In the new context, with the biasing contextual information, most of the fingerprint experts made a no-match decision, thus contradicting their own previous match identification decisions.

Nakhaeizadeh et al. (2014) explored whether forensic anthropology experts would be vulnerable to context effects in the assessment of unidentified skeletal remains. Sure enough, experts exposed to contextual information about the sex, ancestry, and age at death of the

skeletal remains were affected by that information in their interpretation and conclusion of the remains. Compared to experts in a control condition who received no contextual information about the skeletal remains, experts in the contextually biasing information conditions confirmed the extraneous contextual information in their biological profile determinations. For example, only 31% of participants in the control group concluded that the skeletal remains were male. However, among experts in the contextually biasing condition with extraneous information suggesting the remains were male, 72% concluded the remains were male, and among those in the contextually biasing condition with extraneous information suggesting the remains were female, 0% concluded the remains were male. Contextual biases like those described in these studies of forensic scientists are likely present in the work of forensic mental health too.

Judges are experienced, well-trained, and typically highly motivated to be accurate and fair. Nevertheless, clear empirical evidence has been established showing the existence of unintentional cognitive biases affecting the judgments and decisions of judges. For example, Guthrie et al. (2001) found in a sample of 167 federal magistrate judges that they were susceptible to anchoring effects, framing effects, hindsight bias, the representativeness heuristic, and egocentric biases on their judicial decision making. Similarly, Wistrich et al. (2005) showed that judges generally cannot disregard inadmissible information in their legal decisions – even when they were reminded, or they themselves had ruled, that the information was inadmissible. For example, inadmissible information about demands disclosed during settlement conferences, conversations protected by attorney-client privilege, prior sexual history of alleged rape victims, prior criminal convictions of plaintiffs, and information the government had promised not to rely on at sentencing influenced judges' decisions despite their best efforts to ignore the inadmissible

information. Like the contextual biases that have been shown to affect forensic scientists, these various cognitive biases that affect judges do so despite motivation and effort to be unbiased.

Similarly, Helm et al. (2016) found that elite arbitrators are human too, and subject to unintentional System 1-induced biases on their judgments and decisions. Arbitration is an increasingly common type of alternative dispute resolution that provides an alternative to filing a lawsuit and going to court (the traditional method for resolving disputes). Arbitrators resolve thousands of disputes every year, including some high-stakes cases (Helm et al.). Like judges, arbitrators are motivated to be unbiased and fair in their decisions, though judges and arbitrators differ in several ways. Elite arbitrators are highly trained and experienced in specialized areas. Helm et al. studied elite arbitrators specializing in resolving commercial disputes to determine whether these kinds of experts are susceptible to cognitive biases in their work. As might be expected based on the other studies reviewed in this chapter, Helm et al. found that elite arbitrators are subject to the conjunction fallacy, framing effects, confirmation bias, and that their excessive reliance on intuition may exacerbate the effects of System 1 biases on their professional judgments and decisions. These System 1 biases are likely to affect forensic mental health experts too – and they may also excessively rely on intuition in problematic ways.

Similar to the training processes and ethics codes in forensic mental health, professional accountants are trained that objectivity is paramount to their work (Moore et al., 2003). And like in forensic mental health, several sources assume that auditor bias is a matter of deliberate choice – that is, auditors are assumed to be able to complete high-quality, objective audits if they so choose (Bazerman, Loewenstein, & Moore, 2002). However, as Moore and colleagues show, the biases that typically affect professional auditors is pervasive, unconscious, and unintentional. Moore et al. showed through three experiments that auditors' judgments are unintentionally

biased in favor of their clients (the "self-serving bias") and that the bias is not easily corrected because auditors are not fully aware of the bias or how it affects their judgments (despite incentives to be objective).

Moore et al. (2003) and Bazerman et al. (2002) make some important points about the conditions that bias professional auditors that are worth mentioning because similar conditions exist in forensic mental health. Bazerman et al. assert that three structural aspects of accounting create substantial opportunities for bias to influence – two of which in particular highly relevant to forensic psychology. The first is ambiguity. Bias thrives whenever information can be interpreted in different ways – greater ambiguity leads to more biased information processing and outcomes (Kunda, 1990; Thompson & Loewenstein, 1992). Auditors must accumulate and synthesize a great deal of information to make judgements about client firms – just as forensic psychologists must do to make judgements about case referrals. And like auditing, forensic mental health is "art" in addition to some science. The imprecision inherent in auditing and forensic mental health allow motivated reasoning to bias experts judgments. The second condition is attachment, which as we have already described breeds the self-interest bias in accountants (Bazerman et al.; Moore et al.) and forensic mental health experts alike (Murrie et al., 2013).

Social Psychology: Explicit and Implicit Social Cognition and Social-Cognitive Biases

Social psychology is a branch of psychology that focuses on how people affect one another's behaviors, and social cognition is a part of social psychology focused on how other people affect a given person's cognitive functions, such as perceptions, reasoning, attention, memory, and decision making (Devine et al., 1994). Social cognition emerged in the late 1970s (Devine et al.), and dual-process theories in social cognition became increasingly common in the

1980s and 1990s (e.g., Chaiken, 1980; Chaiken & Trope, 1999; Devine, 1989; Greenwald & Banaji, 1995; Petty & Cacioppo, 1986).

Social behavior and judgments were historically assumed to be under conscious control. However, researchers began recognizing that social behavior often is affected by experiences in a manner not known by the actor (Greenwald & Banaji, 1995). Dual-process theories began emerging in social psychology, theorizing that many social-cognitive judgments and behaviors have important implicit (System 1) modes of operation. These dual-process theories in social cognition have attempted to describe implicit, unconscious, heuristic processes (System 1) as separate but certainly related to explicit, conscious, and deliberative processes (System 2) in such areas as attitudes (e.g., Chaiken, 1980; Greenwald & Banaji; Petty & Cacioppo, 1986), perceptions of the self (e.g., Greenwald & Banaji), stereotypes (e.g., Devine, 1989; Greenwald & Banaji), and perception of others (e.g., Chaiken & Trope, 1999), among other areas. In this chapter, we will review the empirical evidence to date on implicit (System 1) processes including attitudes, perceptions of oneself, and stereotypes as they affect forensic mental health experts' judgments and decisions as well as experts in fields who face similar judgment tasks.

Social Psychological Biases and Forensic Mental Health Judgments

People's (including experts') perceptions, judgments, and decisions can be influenced by other people, by their expectations about other people's perceptions, by subtle features of the environment, and by their own pre-existing attitudes and beliefs. Furthermore, people can be biased even when motivated not to be, and when people compare their own biases against others, they have a much harder time seeing their own biases than they do biases in other people. This section examines these kinds of social psychological biases as they might affect forensic mental

health judgments and decisions as well as the judgments and decisions of experts in fields who face similar decision tasks.

For example, we review evidence about how attitudes toward issues such as capital punishment and gender equality unintentionally affect experts' judgments and decisions, how perceptions of oneself as compared to others are unintentionally distorted due to introspective failures (because by definition these implicit processes are below one's level of conscious awareness), and how stereotypes about race and ethnicity unintentionally affect experts' judgments and decisions.

Evidence to date about social-cognitive biases relevant to forensic mental health judgments. Implicit (System 1) social psychological processes have been shown to affect the judgments and decisions of forensic mental health professionals. Even for experts who are motivated to be unbiased, there is mounting evidence that forensic mental health experts are susceptible to System 1-induced social-cognitive biases by virtue of being human. For example, evidence has emerged that forensic psychologists' capital punishment attitudes affect their judgments and decisions in capital cases (Neal, 2016; Neal & Cramer, under review). And an insidious social-cognitive bias called the "bias blind spot," which refers to an exceedingly common human tendency to recognize bias in others but fail to recognize it in oneself (Pronin, Lin, & Ross, 2002), has been documented in forensic mental health professionals (Commons, Miller, & Gutheil, 2004; Neal & Brodsky, 2016; Neal et al., 2017; Zapf, Kukucka, Kassin, & Dror, 2017). Although we sought to review research on how stereotypes affect forensic mental health professionals' judgments, we could locate no research on this issue to date.

Attitudes. Neal (2016) sought to answer whether forensic mental health experts' preexisting attitudes might affect their professional judgments and decisions. To investigate this

question, Neal focused on death penalty attitudes and decisions relevant to capital cases. She measured the death penalty attitudes of 206 forensic psychologists (using the Death Penalty Attitudes Scale, O'Neil, Patry, & Penrod, 2004) and asked the forensic psychologist respondents whether they would work for the prosecution, defense, and/or court in capital cases. She hypothesized that evaluator attitudes toward capital punishment would systematically influence their willingness to accept capital case referrals, such that evaluators with strong support would be more likely to work for the prosecution and evaluators with low support would be more likely to work for the defense. These hypotheses were partially supported.

As hypothesized, lower support was associated with being willing to work for the defense, as well as a higher likelihood of rejecting any referral from any source (abstaining completely from capital case evaluations). And stronger support was associated with higher willingness to be involved in capital cases across any referral source. No psychologists reported selective willingness to work only for the prosecution (though several did report selective willingness to work only for the defense – correlated with the strength of their opposition to the death penalty). Neal asserted these findings raise the specter of systematically biased involvement of forensic psychologists in capital case evaluations based on their death penalty attitudes. She also suggested these findings provide a partial explanation for the "allegiance effect" such that evaluators' preexisting attitudes may influence their selective participation in the legal process via "filtering" effects. Future research is needed to further explore the effect of experts' preexisting attitudes and whether these attitudes transfer to biased decision-making in the cases themselves. In this study, attitudes were measured explicitly rather than implicitly – future work must measure the effects of implicit attitudes on judgments.

Neal and Cramer (in progress) studied forensic psychologists death penalty attitudes again, this time studying whether these attitudes were systematically related to forensic psychologists' willingness to conduct the most ethically questionable clinical task in the criminal justice system: competence for execution evaluations. Although there was no direct effect of death penalty attitudes on willingness to accept competence for execution referrals, that relationship was fully mediated by moral disengagement. Moral disengagement is a social-cognitive process through which people reason their way toward harming others (Bandura, 2015). Thus, moral disengagement served as a theoretical "bridge" between forensic psychologists' attitudes and judgments, an interesting finding for helping clarify how psychologists decide to engage in competence for execution evaluations. Here again, the measures were explicit rather than implicit, and studies of implicit attitudes are needed.

The Self: Bias Blind Spot. The bias blind spot has been found across many social groups: college students believe they are less biased than their fellow students, airline passengers think they are less biased than other passengers, car drivers on average believe themselves to be above-average drivers, and so forth. Pronin and colleagues (2002) conducted a series of studies looking at multiple biases by having participants self-report various biases, and then indicate how much the average American was biased. Participants overwhelmingly reported that they personally were less biased than the average American across many different types of biasing situations, showing the generalizability of this concept.

The bias blind spot is theorized to arise from the interplay of two phenomena: the introspection illusion and the naive realism. People tend to self-evaluate the extent of bias in their own behavior through introspection. Since introspection is unlikely to reveal biased thought processes (due to these implicit processes occurring below the level of conscious awareness),

they typically go unnoticed and uncorrected. Naive realism is the conviction that oneself interacts with the world objectively and therefore one's behavior sufficiently reflects a rational response to the environment, while others' respond in ways that are not grounded in reality (Pronin, Gilovich & Ross, 2004, Scopelliti et al., 2015).

Commons et al. (2004) asked forensic psychiatrists to rate their potential bias in a number of situations, including recent cases in which they had served as an expert witness. They concluded that forensic psychiatrists markedly underestimate their own biases compared to their peers, consistent with the bias blind spot. Moreover, some situations were perceived as more biasing than others, and participants underestimated the biasing effects of conflicts of interest for both themselves and opposing experts.

Neal and Brodsky (2016) found evidence suggestive of a bias blind spot in forensic psychologists. Using deep narrative interviews in a qualitative study with board-certified forensic psychologists, they found that forensic evaluators perceived themselves as less vulnerable to bias compared to their colleagues. Participants had no trouble identifying bias in their colleagues (100% of the sample discussed ways in which they had observed bias in their colleagues), but many fewer reported any concern about bias in themselves (60% mentioned any concern). Some even reported that they take over cases that might be considered a challenge for others because they believe they are able to control themselves in a way that their colleagues cannot. This finding was theorized to be a result of bias blind spot-induced overconfidence in one's own judgments, a negative consequence that could lead to risky decision making and rejecting aid that may reduce bias and improve validity.

In a recent study of 1,099 forensic mental health professionals across 39 different countries, Zapf et al. (2017) again found evidence of the bias blind spot. In their sample, the

mean accuracy rating that forensic mental health professionals provided for forensic evaluations was 78.5%, yet the experts estimated their own accuracy in forensic evaluations at 81.85% ($p <$ 0.001, $d = .59$ [a medium effect size]). Furthermore, they rated other forensic mental health professionals as more susceptible to cognitive bias in their forensic evaluations (78.1%) than themselves (52.2%).

In the Neal et al. (2017) study of confirmation bias in forensic mental health diagnostic judgments described earlier in this chapter, participants were also asked to rate the extent to which their own forensic work is influenced by bias ($M = 4.74$, $SD = 1.95$) as well as the extent to which work by other forensic psychologists is influenced by bias ($M = 5.19$, $SD = 1.86$), each on a 1 (never) to 9 (always) point Likert-scale. These questions were designed to measure the bias blind spot, the tendency to deny personal bias even while recognizing it in others. They calculated the size of an expert's bias blind spot by subtracting self-rating from other-rating ($M =$ 0.51, $SD = 1.01$). They hypothesized the size of participants' bias blind spot would be positively related to confirmatory bias. They also hypothesized that cognitive reflection tendencies (the ability to resist incorrect heuristic responses in favor of deliberative thought) would be inversely related to the size of the bias blind spot.

The bias blind spot hypothesis yielded a meaningful effect size (i.e., theoretically significant) in the predicted direction, but was not statistically significant. Each additional unit of discrepancy between self and other ratings of bias (i.e., increasing bias blind spot) more than doubled the odds of forensic clinicians engaging in confirmation bias, $B = 0.71$, $Wald(1) = 2.10$, $p = 0.147$, $Exp(B) = 2.02$ (logistic regression model $\chi^2 [1] = 2.33$, $p = 0.127$). The prediction that cognitive reflection tendencies would be inversely related to the bias blind spot emerged with a small effect size in the predicted direction, but the trend did not reach statistical significance.

Forensic clinicians with higher cognitive reflection tendencies had somewhat smaller bias blind spots, $r = -0.176$, $p = 0.092$.

Related literature on social-cognitive biases in similar types of judgment tasks. Outside of the forensic mental health contexts, the influence of implicit (System 1) attitudes, self-perceptions, and stereotypes on related types of expert judgment tasks are reviewed here. These findings are relevant to forensic mental health, because it is likely that many of these same implicit social-psychological processes affect the judgments of forensic mental health professionals in similar ways. For example, professional arbitrators' (Girvan, Deason, and Borgida (2015) and law students' (Braman & Nelson, 2007) legal decisions are systematically affected by their attitudes, the bias blind spot has been documented in forensic scientists (Kukucka, Kassin, Dror, & Zapf, 2017), and judges and physicians show evidence of implicit racial biases in their legal and medical decisions (Green et al., 2007; Rachlinski, Johnson, Wistrich, & Guthrie, 2009).

Attitudes. In a study with professional arbitrators who work in the area of labor arbitration, Girvan et al. (2015) measured the experts' explicit and implicit gender attitudes. In the first part of their study, an experimental lab-based study, they did not find evidence that the experts gender attitudes affected their decisions in two mock arbitration cases in which the gender of the employee-grievants were manipulated (though non-expert undergraduate students did show the expected gender biases in their decisions). However, in a second study, arbitrators' explicit and implicit gender attitudes did predict their decisions in actual published labor arbitration cases. Girvan et al. concluded that implicit and explicit attitudes are important to understand as they affect experts' legal decisions, but that laboratory experiments may not

capture the nature of these attitudes' effects on real decisions – an important insight for continued research in this area.

The effect of attitudes about civil rights (gay rights in this case) on legally-relevant perceptions and judgments by law students was studied by Braman and Nelson (2007). These researchers had participants (both law students and undergraduate students) participate in an experiment in which they studied motivated reasoning in legal judgments (see Kunda, 1990). Specifically, they asked participants to make legally-relevant decisions about the similarity and applicability of previous case law on the target case in the study. In the law, legal precedent (previous case decisions) are used as a guide for deciding future similar cases. But judges have some flexibility in determining what cases are most similar and whether or not a given case is analogous enough to a target case to cite as legal precedent.

The target case used in Braman and Nelson's (2007) study was the *Boy Scouts of the United States of American v. Dale* (2000) case in which a gay male claimed unlawful discrimination by the Boy Scouts after having been removed as a youth leader and dismissed from the organization based on his sexual orientation. The researchers varied some aspects of the case in an experimental design, including the outcome of source cases, such that some participants saw a source case (potential legal precedent) in which unlawful discrimination had been found, or a source case in which the defendant had been found to be acting within his rights. Participants were asked to rate how similar the source case was to the target case on a four-point scale (i.e., how analogous the source case was to cite as legal precedent for deciding the target case). They also measured participant's support on a six-point scale on a question about how acceptable it is for a gay man to serve as a Boy Scout.

Results revealed that attitudes influenced their legal judgments by affecting the perceived similarity between the source cases and the target case. Participants found the source cases that supported their policy views in the target case as more relevant to that litigation. Braman and Nelson (2007) found that legal training did not seem to attenuate motivated perceptions: the law students' attitudes influenced their judgments just as undergraduate students' attitudes influenced their judgments.

This question about whether legal training attenuates bias is more complicated than this, however. Kahan, Hoffman, Evans, Devins, Lucci, & Cheng (2016) found that legal training did protect judges' and lawyers' statutory interpretations from their cultural values compared to general public participants making the same legal interpretations. They noted that law students' interpretations, however, were somewhat biased by their cultural values. Girvan (2016) showed through an elegant experimental design that not all legal professionals' judgments are protected from bias by virtue of their legal training. He showed that learning and applying legal rules (i.e., removes discretion) does protect legal experts from stereotype-induced bias compared to novices, but that learning and applying legal standards (i.e., discretion allowed) does not protect legal experts.

The Self: Bias Blind Spot. Similar to findings that have emerged in forensic mental health, recent findings of the bias blind spot in forensic scientists' perceptions of themselves has also emerged. Kukucka et al. (2017) conducted a recent study of 403 forensic scientists across domains (e.g., latent prints, questioned documents, toxicology, biology/serology/DNA, crime scene investigation, bitemarks, and firearms/toolmarks) from 21 countries. The mean accuracy rating that forensic scientists provided for their field was 94.41%, yet the experts estimated their own accuracy at 96.25% ($p < 0.001$, $d = 0.27$ [a small effect size]). Furthermore, consistent with

the bias blind spot, they rated other forensic scientists as more susceptible to cognitive bias in their work (70.1%) than themselves (25.7%).

Stereotypes. Although we could not locate any studies of how stereotypes affect forensic mental health professionals' judgments, there are provocative studies of how implicit racial biases affect judicial (Rachlinski et al., 2009) and physician (Green et al., 2007) decision making that is likely relevant for forensic psychology. In both of these studies, the researchers used a tool called the Implicit Association Test (IAT) focused on race to measure participants' implicit attitudes about race. The IAT was developed through decades of research on bias and stereotypes, and is typically administered via computer through a sorting task in which participants pair words and faces. These researchers found that experts typically do not endorse or exhibit explicit (System 2) stereotype-based biases that affect their professional judgments and decisions. But in both studies, implicit (System 1) stereotype-based biases unintentionally affected the judges' and doctors' judgments and decisions.

Rachlinski et al. (2009) asked judge participants to complete the computer-based IAT first, and then respond to hypothetical vignettes and make decisions about them. In some vignettes, the race of the defendant was subliminally primed but not identified explicitly, whereas in the final vignette the race of the defendant was made explicit. Results from this study, first, showed that judges hold implicit racial biases that mirror the general public's implicit attitudes about race (in this IAT's case, a systematic white over black preference). Second, the results showed that judges are not explicitly biased: even when primed with race-related primes, those primes did not directly affect their judgments.

However, the implicit racial attitudes of judges did affect their legal judgments (Rachlinski et al., 2009). Judges' scores on the race IAT had a consistent, marginally significant

influence on their judgments across two different vignettes. Judges with stronger white preference on the IAT gave harsher sentences to defendants when they had been primed with black-associated words (e.g., graffiti, Harlem, homeboy, jerricurl, rap, segregation, basketball, gospel, afro, reggae, athlete) compared to judges who were primed with neutral words (e.g., baby, heaven, coffin, summer, truth, accident, mosquito, virus, toothache, rainbow, paralysis). And judges who exhibited a black preference on the IAT gave less harsh sentences to defendants when they had been primed with black-associated words compared to judges primed with neutral words.

Green and colleagues (2007) studied how implicit racial biases affect physicians' clinical decision making. They asked physician-participants to respond to a clinical vignette of a patient presenting to an emergency room with an acute heart problem and to make a decision about whether or not to use thrombolysis, a treatment to dissolve blood clots. Physician participants were randomly assigned to either a black or white patient vignette. They were asked to complete the race IAT to measure their implicit racial attitudes as well as respond to questions about perceptions of patient cooperativeness, attribution of symptoms to coronary artery disease, and respond to a questionnaire about their explicit racial attitudes.

Results indicated that physicians did not endorse explicitly racial attitudes (Green et al., 2007). However, like judges and the general public, physicians exhibited an implicit racial preference favoring white people over black people, and they endorsed implicit stereotypes of black people as less cooperative with medical procedures and generally. Furthermore, these implicit attitudes affected their clinical decision making. As the strength of physicians' implicit prowhite bias increased, their likelihood of treating white patients with thrombolysis and not treating black patients with thrombolysis increased. These results demonstration the

disconnection between implicit and explicit attitudes, and the predictive validity of implicit attitudes for experts' judgments. Results suggest physicians' implicit racial biases may contribute to the minority health disparities present in this country.

Conclusion

In this chapter, we have reviewed evidence of unintentional biases in experts' judgments. We have focused especially on forensic mental health experts, but we've also described relevant findings from other professional arenas in which experts face similar decision tasks. We started with foundational background from cognitive and social psychology, explaining dual-process theories of cognition in both subfields of psychology to ground our discussion of bias. By "bias" in this chapter, we focus exclusively on System 1-induced cognitive and social-cognitive biases. These are unintentional, understandable, predictable, systematic errors that experts make by virtue of being human. When we understand the contexts and situations in which these types of biases emerge, we can work to change these contexts and situations so as to minimize the likelihood of biases exerting negative effects on experts' judgments. In reviewing this literature, we have identified several gaps that future research can address.

Future directions for research. Most of the empirical evidence of cognitive biases in experts' judgments, as well as the evidence of social-cognitive biases in experts' judgments, are outside of the forensic mental health domain. We did review the growing and pretty new body of evidence regarding cognitive biases in forensic mental health judgments, but a lot of this work is not yet published and has not gone through the peer-review system. Most of the evidence about how cognitive biases influence experts' judgments is in other areas (e.g., forensic science, law). Thus, there is a need for researchers to conduct methodologically strong, ecologically valid experimental research in forensic mental health domains to further clarify how and when

cognitive biases affect forensic mental health experts' judgments and decisions. When we better understand how cognitive biases work in this domain, we can better design mitigation procedures.

Similarly, very little of the evidence on implicit social-cognitive biases is in the domain of forensic mental health. In fact, only emerging information about the "bias blind spot" is in the forensic mental health domain. We did review some studies of how forensic mental health professionals' attitudes influence their judgments; however, these attitudes have only been measured explicitly to date and it is not clear whether implicit attitudes also affect forensic mental health judgments. It is highly likely, given findings from law (e.g., Girvan et al., 2015), but as of yet remains unstudied in forensic mental health.

We were unable to find a single study of how implicit stereotypes affect the judgments and decisions of forensic mental health experts. There is compelling evidence from law (Rachlinski et al., 2009) and medicine (Green et al., 2007) that implicit racial stereotypes will likely influence forensic mental health professionals' judgments too, but as of yet there appear to be no studies of this issue. Furthermore, there is a need for studying how other kinds of implicit stereotypes (not just race) affect professionals' decisions – both in forensic mental health as well as in other domains. This area is ripe for future research.

In sum, there is strong theoretical reasons to suspect that forensic mental health professionals are subject to the effects of unintentional System 1-induced cognitive and social-cognitive biases in their judgments and decisions. Emerging research from the forensic mental health domain supports these theoretical predictions, but much work remains to be done to better understand the boundary conditions of when and how these particular kinds of experts are biased in the particular environmental contexts in which they work. As more research in this area

emerges, it will flesh out the unique ways in which the contextual environments of forensic mental health affect our judgments. It will also inform what can be done to mitigate the effects of these unintentional and unwanted biases in forensic mental health professionals' judgments and decisions.

References

Arkes, H.R., Wortmann, R.L., Saville, P.D., & Harkness, A.R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology, 66,* 252–254.

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling, 2*, 43-63.

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives*, *11*, 109-126.

Bandura, A. (2015). *Moral disengagement: How people do harm and live with themselves*. New York: Worth Publishing.

Bazerman, M.H., Loewenstein, G., & Moore, D.A. (2002). Why good accountants do bad audits. *Harvard Business Review, 80*, 96-103.

Beltrani, A., Reed, A., Zapf, P. A., Dror, I. E., & Otto, R. K. (2017, March). *Is hindsight really 20/20? The impact of outcomes on the decision making process.* Paper presented at the annual conference of the American Psychology-Law Society, Seattle, WA.

Braman, E., & Nelson, T. E. (2007). Mechanism of motivated reasoning? Analogical perception in discrimination disputes. *American Journal of Political Science*, *51*, 940-956. doi: 10.1111/j.1540-5907.2007.00290.x

Boy Scouts of the United States of American v. Dale, 530 US 640 (2000).

Caplan, R.A., Posner K.L., & Cheney, F.W. (1991). Effect of outcome on physician judgments of appropriateness of care. *Journal of the American Medical Association, 265*, 1957–1960.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology, 39*, 752-766. doi: 10.1037/0022-3514.39.5.752

Chaiken, S. & Trope, Y. (Eds) (1999). *Dual-process theories in social psychology.* New York: Guilford.

Commons, M. L., Miller, P. M., & Gutheil, T. G. (2004). Expert witness perceptions of bias in experts. *The Journal of the American Academy of Psychiatry and the Law*, *32*(1), 70–75.

Cook, M.B., & Smallman, H.S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human factors: The journal of the human factors and ergonomics society*, *50*, 745-754.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18. doi: 10.1037/0022-3514.56.1.5

Devine, P.G., Hamilton, D.L.E., & Ostrom, T.M. (Eds). (1994). *Social cognition: Impact on social psychology.* Cambridge, MA: Academic Press.

Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again sustained inattentional blindness in expert observers. *Psychological science*, *24*, 1848-1853.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*, 74-78. doi:10.1016/j.forsciint.2005.10.017

Epperson, D. L., Kaul, J. D., Goldman, R., Hout, S., Hesselton, D., & Alexander, W. (1998). *Minnesota Sex Offender Screening Tool—Revised (MnSOST-R).* St. Paul: Minnesota Department of Corrections. Available online at http://www.psychology .iastate.edu

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. doi: 10.1146/annurev.psych.59.103006.093629

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, *1*, 288-299. doi: 10.1037/0096-1523.1.3.288

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*, 25–42. http://doi.org/10.1257/089533005775196732

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases." *European Review of Social Psychology*, *2*, 83–115. http://doi.org/10.1080/14792779143000033.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*, 592–596. http://doi.org/10.1037/0033-295X.103.3.592

Gigerenzer, G. (2008). Work Why Heuristics. *Perspectives on Psychological Science*, *3*, 20–29.

Girvan, E. J. (2016). Wise restraints?: Learning legal rules, not standards, reduces the effects of stereotypes in legal decision-making. *Psychology, Public Policy, and Law*, *22*, 31–45. doi:10.1037/law0000068

Girvan, E.J., Deason, G., & Borgida, E. (2015). The generalizability of gender bias: Testing the effects of contextual, explicit, and implicit sexism on labor arbitration decisions. *Law and Human Behavior*, *39*, 525-537. doi: 10.1037/lhb0000139

Green, A.R., Carney, D.R., Pallin, D.J., Ngo, L.H., Raymond, K.L., Iezzoni, L.I., et al. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine, 22,* 1231-1238. doi: 10.1007/s11606-007-0258-5

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4-27. doi: 10.1037/0033-295X.102.1.4

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *Cornell Law Review, 86,* 777-778.

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders (User Report 99–02).* Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hare, R. D. (2003). *The Hare Psychopathy Checklist—Revised (2nd ed.).* Toronto, Ontario, Canada: Multi-Health Systems.

Haselton, M., Bryant, G.A., Wilke, A., Frederick, D., & Galperin, A. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition, 27,* 733-763.

Hastdorf, A.H. & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49,* 129-134.

Helm, R. K., Wistrich, A. J., & Rachlinski, J. J. (2016). Are Arbitrators Human? *Journal of Empirical Legal Studies, 13,* 666-692. doi: 10.1111/jels.12129

Kahan, D. M., Hoffman, D., Evans, D., Devins, N., Lucci, E., & Cheng, K. (2016). 'Ideology' or 'situation sense'? An experimental investigation of motivated reasoning and professional judgment. *University of Pennsylvania Law Review,* 2, 394.

Kahneman, D. (2003). A perspective on judgment and choice - Mapping bounded rationality. *American Psychologist, 58,* 697–720. http://doi.org/10.1037/0003-066x.58.9.697.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64,* 515-526. doi: 10.1037/a0016755

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454. http://doi.org/10.1016/0010-0285(72)90016-3.

Kukucka, J., Kassin, S.M., Dror, I.E., & Zapf, P.A. (2017, March). *Cognitive bias: A survey of forensic examiners*. Paper presented at the annual conference of the American Psychology-Law Society, Seattle, WA.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498. doi:10.1037/0033-2909.108.3.480

LaBine, S.J. & LaBine, G. (1996). Determinations of negligence and the hindsight bias. *Law and Human Behavior, 20,* 501–516.

LeBourgeois III, H. W., Pinals, D. A., Williams, V., & Appelbaum, P. S. (2007). Hindsight bias among psychiatrists. *Journal of the American Academy of Psychiatry and the Law, 35,* 67-73.

McAuliff, B. D., & Arter, J. L. (2016). Adversarial allegiance: The devil is in the evidence details, not just on the witness stand. *Law and Human Behavior, 40,* 524–535. doi:10.1037/lhb0000198

Moore, D.A., Loewenstein, G., Tanlu, L., & Bazerman, M.H. (2003). Auditor independence, conflict of interest, and the unconscious intrusion of bias. *Harvard Business School Working Paper No 03-116.* Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.2829&rep=rep1&type=pdf

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science, 24,* 1889–1897. doi:10.1177/0956797613481812

Murrie, D.C., Boccaccini, M., Johnson, J., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy Checklist scores in Sexually Violent Predator trials suggest partisan allegiance in forensic evaluation? *Law and Human Behavior, 32,* 352-362. doi: 10.1007/s10979-007-9097-5

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15,* 19–53. doi:10.1037/a0014897

Nakhaeizadeh, S., Dror, I.E., & Morgan, R.M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science and Justice, 54,* 208-214. doi: 10.1016/j.scijus.2013.11.003

Neal, T. M. S. (2016). Are Forensic Experts Already Biased before Adversarial Legal Parties Hire Them? *PLoS ONE*, *11*(4), e0154434. doi:10.1371/journal.pone.0154434

Neal, T. M. S., & Brodsky, S. L. (2016). Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *22*, 58–76. doi:10.1037/law0000077

Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law, 20*, 200–211. doi:10.1037/a0035824

Neal, T.M.S., MacLean, N., Morgan, R.D., & Murrie, D.C. (2017, March). *Robust evidence of confirmation bias in forensic psychologists' diagnostic reasoning.* Paper presented at the annual conference of the American Psychology-Law Society, Seattle, WA.

Neal, T.M.S. & Saks, M.J. (in preparation). Context effects in forensic mental health science: A review and application of the science of science to the practice of forensic mental health evaluations.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

O'Neil K.M., Patry M.W., & Penrod, S.D. (2004). Exploring the effects of attitudes toward the death penalty on capital sentencing verdicts. *Psychology, Public Policy, and Law, 10*, 443–470.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology, 19*, 123-205.

Popper, K. R. (1959). The logic of scientific discovery. *London: Hutchinson*.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111, 781-799. doi: 11.1037/0033-295X.111.3.781

Pronin, E., Lin, D. L., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus other. *Personality and Social Psychology Bulletin, 28*, 369-381. doi: 10.1177/0146167202286008

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review, 84*, 1195-1246.

Sacchi, S. & Cherubini, P. (2004). The effect of outcome information on doctors' evaluations of their own diagnostic decisions. *Medical Education, 38*, 1025–1027.

Scopelliti, I., Morewedge, C.K., McCormick, E., Min, L. H., Lebrecht, S., & Kassam, K. S. (2015). Bias blind spot: Structure, measurement, and consequences. *Management Science*, 61, 2468-2486.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22. http://doi.org/10.1037/0033-2909.119.1.3.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research, 177*, 1333–1352. http://doi.org/10.1016/j.ejor.2005.04.006

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? *The Behavioral and Brain Sciences, 23*, 645-665-726. http://doi.org/10.1017/S0140525X00003435.

Thompson, L. & Loewenstein, G. (1992). Egocentric interpretation of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes, 51*, 176-197.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology, 5*, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131. doi:10.1126/science.185.4157.1124

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 273–281. doi:10.1080/14640746808400161

Wistrich, A. J., Guthrie, C., & Rachlinski, J. J. (2005). Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *University of Pennsylvania Law Review, 153*, 1251-1345.

Zapf, P.A., Kukucka, J., Kassin, S.M., & Dror, I.E. (2017, March). *Cognitive bias: A survey of forensic evaluators.* Paper presented at the annual conference of the American Psychology-Law Society, Seattle, WA.